

Bayesian data assimilation based on a family of outer measures

Jeremie Houssineau and Daniel E. Clark

Abstract—A flexible representation of uncertainty that remains within the standard framework of probabilistic measure theory is presented along with a study of its properties. This representation relies on a specific type of outer measure that is based on the measure of a supremum, hence combining additive and highly sub-additive components. It is shown that this type of outer measure enables the introduction of intuitive concepts such as pullback and general data assimilation operations.

Index Terms—Outer measure; Data assimilation

INTRODUCTION

Uncertainty can be considered as being inherent to all types of knowledge about any given physical system and needs to be quantified in order to obtain meaningful characteristics for the considered system. Uncertainty originates from randomness as well as from lack of precise knowledge and is usually modelled by a probability measure on a state space on which the system of interest can be represented. However, in situations where the lack of knowledge is acute, probability measures are not always suitable for representing the associated uncertainty, and different approaches have been introduced based on *fuzzy sets* [33], *upper and lower probabilities* [10] or *possibilities* and *plausibilities* [28]. Various extensions of the concept of probability have also been studied [32, 18].

In this article, we study a representation of uncertainty that is completely within the standard framework of probabilistic measure theory. We do not claim that the approach considered here is more general than any previously introduced representation of uncertainty, it is rather the considered framework that will prove to be beneficial when handling uncertainty. Indeed, the proposed approach allows for deriving a Bayesian data assimilation operation that combines independent sources of information. It is shown that this operation yields Dempster’s rule of combination [11, 29] when the proposed representation of uncertainty reduces to a plausibility, which furthers the relation between the Bayesian and Dempster-Shafer approaches discussed in [5]. The mathematical properties of the proposed Bayesian data assimilation operation are also studied, similarly to [5]. This operation can be used in conjunction with random variables describing different types of physical systems, for instance it has been used in [19, Chapt. 2] together with a general representation of stochastic populations [20] in order to derive a principled solution to the *data association* problem which underlies data assimilation for probabilistic multi-object dynamical systems and which is usually handled in an ad-hoc

way. This enabled new methods for multi-target filtering to be derived, as in [9] and [19, Chapt. 4].

The proposed representation of uncertainty also enables the definition of a meaningful notion of *pullback* – as the inverse of the usual concept of pushforward measure – that does not require a modification of the measurable space on which the associated function is defined, i.e. it does not require the considered measurable function to be made bi-measurable.

The structure of the paper is as follows. Examples of limitations encountered when using probability measures to represent uncertainty are given in Section I. The concepts of outer measure and of convolution of measures are recalled in Section II, followed by the introduction of the proposed representation of uncertainty in Section III and examples of use in Section IV. Preliminary results about pushforward and pullback measures are given in Section V before the statement of the main theorems in Section VI.

Notations

The sets of non-negative integers is denoted \mathbb{N} , and the set of real (non-negative) numbers is denoted \mathbb{R} (\mathbb{R}^+). The power set of a set A is denoted $\wp(A)$, where “ \wp ” is referred to as the “Weierstrass p ”. For any set E , the function on E which is everywhere equal to one is denoted $\mathbf{1}$. For any set F , any $f, f' : E \rightarrow F$ and any $g, g' : E \rightarrow \mathbb{R}$, we define the mappings $g \cdot g'$ and $f \times f'$ on E as

$$\begin{aligned} g \cdot g' : x &\mapsto g(x)g'(x) \in \mathbb{R}, \\ f \times f' : x &\mapsto (f(x), f'(x)) \in F \times F, \end{aligned}$$

and the mappings $g \rtimes g'$ and $f \rtimes f'$ on $E \times E$ as

$$\begin{aligned} g \rtimes g' : (x, x') &\mapsto g(x)g'(x') \in \mathbb{R}, \\ f \rtimes f' : (x, x') &\mapsto (f(x), f'(x')) \in F \times F. \end{aligned}$$

We consider two complete probability spaces $(\Omega, \Sigma, \mathbb{P})$ and $(\Omega', \Sigma', \mathbb{P}')$ that are assumed to represent independent sources of uncertainty. If two random variables X and X' are defined on Ω , then the associated joint random variable is $X \rtimes X'$ and its law is found to be the pushforward $(X \rtimes X')_* \mathbb{P}$, moreover, if X and X' are real random variables then their product is $X \cdot X'$. If X' is defined on Ω' then the corresponding joint random variable will be $X \rtimes X'$ and its law will be $(X \rtimes X')_*(\mathbb{P} \rtimes \mathbb{P}')$.

If (E, \mathcal{E}) is a measurable space then $\mathbf{M}(E)$ and $\mathbf{M}_1(E)$ denote the set of measures and the set of probability measures on E respectively. If \mathbf{X} is a Polish space, then $\mathcal{B}(\mathbf{X})$ denote the Borel σ -algebra on \mathbf{X} .

J. Houssineau is with the Department of Statistics and Applied Probability, National University of Singapore.

D.E. Clark is with the School of Engineering and Physical Sciences, Heriot-Watt University, Edinburgh.

I. UNCERTAINTY AND PROBABILITY MEASURES

Random experiments are mathematically defined by probability spaces, defining possible events and attributing them a probability. However, the use of a probability space for representing uncertainty that is not induced by pure randomness, i.e. for describing random or non-random experiments about which only partial knowledge is available, can become less intuitive as well as technically challenging in some cases. In the following examples, we study when uncertainty can be encoded into a probability measure as well as cases where the existence of a procedure for doing so becomes less clear.

Example 1. Let X be a random variable from $(\Omega, \Sigma, \mathbb{P})$ to $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ and assume that it is only known that X has its image in the Borel subset A in $\mathcal{B}(\mathbb{R})$ with probability $\alpha \in [0, 1]$. This information can be encoded via the sub- σ -algebra $\mathcal{A} \doteq \{\emptyset, A, A^c, \mathbb{R}\}$ of \mathbb{R} by defining the law p of X on $(\mathbb{R}, \mathcal{A})$ rather than on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ and by setting $p(A) = \alpha$. Similarly, if nothing is known about X , then this can be encoded via the *trivial* sub- σ -algebra $\{\emptyset, \mathbb{R}\}$.

The concept of sub- σ -algebra is useful for describing different levels of knowledge, such as when used for conditional expectations [22, Chapt. 27]. However, we will see in the next example that their use can become challenging in some situations.

Example 2. Let p be a probability measure on $(\mathbf{E}, \mathcal{E})$ and let p' be another probability measure on $(\mathbf{E}, \mathcal{E}')$, with $\mathcal{E}' \subset \mathcal{E}$, then for any scalar $a \in (0, 1)$, the probability measure $q_a = (1 - a)p + ap'$ can only be defined on the coarsest σ -algebra, that is \mathcal{E}' . When considering the extreme case where \mathcal{E}' is the trivial σ -algebra $\{\emptyset, \mathbf{E}\}$, it results that nothing is known about q_a , however small is a . One way to bypass this drawback is to single out a finite reference measure λ in $\mathbf{M}(\mathbf{E})$ and to define an extended version of p' denoted \bar{p}' as a *uniform* probability measure as follows

$$(\forall A \in \mathcal{E}) \quad \bar{p}'(A) \doteq \frac{\lambda(A)}{\lambda(\mathbf{E})}.$$

In this way, the probability of a given event in \mathcal{E} is equal to the probability of any other event of the same “size” with respect to the measure λ . In other words, no area of the space \mathbf{E} is preferred over any other. Besides the facts that a reference measure is required and that the size of the space is limited (there is no uniform measure over the whole real line with respect to the Lebesgue measure), this way of modelling the information is not completely equivalent to the absence of information. There exist ways of modelling uncertainty on a probability measure itself, such as with Dirichlet processes [16] and to some extent with Wishart distributions [31]; yet, these solutions do not directly help with the non-informative case since they require additional parameters to be set up.

Example 3. Consider a discrete-time joint process $(X_n, Y_n)_{n \geq 0}$ assumed to be a hidden Markov model [2] and consider the case where realisations y_n of the observation process are not received directly but are instead known to be in some subset A_n of the observation space. This is in fact what happens in practice since most real

sensors are finite-resolution, e.g. if the sensor is a camera then A_n would be one of the pixels in the received image at time step n . This type of data could be treated as such using non-linear filtering methods, as in [21] for finite-resolution radars; however the associated uncertainty is often modelled by a Gaussian distribution so that the Kalman filter can be used instead. Yet, replacing information of the form $y_n \in A_n$ by a probability distribution can have a great impact on the value of the denominator in Bayes’ theorem. Although this is unimportant in Kalman filtering since the numerator is equally affected, the value of the denominator does matter in different contexts, such as in multi-target tracking [3] where it is used as a weighting coefficient for the relative importance of a given *track* when compared to others [26]. We will see in Section VI that the Gaussian distribution can be replaced by another object that better represents data of the form $y_n \in A_n$ while preserving the advantages of Kalman filtering.

Another important aspect of probability theory is the gap between probability measures on countable and uncountable sets, as explained in the following example.

Example 4. Let \mathbf{X} be a countable set equipped with its discrete σ -algebra and assume that some physical system can be uniquely characterised by its state in \mathbf{X} . Let X be a random variable on $(\Omega, \Sigma, \mathbb{P})$ and X' be another random variable on another probability space $(\Omega', \Sigma', \mathbb{P}')$ and assume that both X and X' represent some uncertainty about the same physical system. The fact that the two random variables represent the same system can be formalised by the event $\Delta \doteq \{(x, x) \text{ s.t. } x \in \mathbf{X}\}$, which is the diagonal of $\mathbf{X} \times \mathbf{X}$. The information contained in the laws $p \doteq X_* \mathbb{P}$ and $p' \doteq (X')_* \mathbb{P}'$ can then be *combined* into a conditional probability measure $\hat{p}(\cdot | \Delta) \in \mathbf{M}_1(\mathbf{X})$, characterised by

$$(\forall B \subseteq \mathbf{X}) \quad \hat{p}(B | \Delta) \doteq \frac{p \rtimes p'(B \times B \cap \Delta)}{p \rtimes p'(\Delta)},$$

where p and p' are assumed to be *compatible*, i.e. that $p \rtimes p'(\Delta) \neq 0$. This result is justified by the transformation of the conditional probability measure $p \rtimes p'(\cdot | \Delta)$ with support on Δ into a probability measure on \mathbf{X} (isomorphism mod 0). Let $w, w' : \mathbf{X} \rightarrow [0, 1]$ be the probability mass functions induced by p and p' and characterised by

$$p = \sum_{x \in \mathbf{X}} w(x) \delta_x, \quad \text{and} \quad p' = \sum_{x \in \mathbf{X}} w'(x) \delta_x,$$

then the probability measure $\hat{p}(\cdot | \Delta)$ can be more naturally characterised via its probability mass function \hat{w} on \mathbf{X} , which is found to be

$$\hat{w} : x \mapsto \frac{w(x)w'(x)}{\sum_{y \in \mathbf{X}} w(y)w'(y)},$$

However, if \mathbf{X} is uncountable and equipped with its Borel σ -algebra $\mathcal{B}(\mathbf{X})$ and if the probability measures p and p' are diffuse on \mathbf{X} , then they will not be compatible by construction. Indeed, even though the diagonal Δ can still be defined and is measurable under extremely weak conditions on \mathbf{X} , it holds that $p \rtimes p'(\Delta) = 0$ (more specifically, the diagonal Δ of $\mathbf{X} \times \mathbf{X}$

is measurable in a separable metric space [4, Lemma 6.4.2], and remains measurable if \mathbf{X} is generalised to a Hausdorff topological space with a countable base; an interesting result from [24], detailed in [27, Chapt. 21], is that the diagonal Δ is never measurable when the cardinality of \mathbf{X} is strictly larger than the cardinality of the continuum.) The fact that $p \times p'(\Delta) = 0$ is caused by the strong assumption that the probabilities $p(B)$ and $p'(B)$ are known for all measurable subsets in $\mathcal{B}(\mathbf{X})$ and, because p and p' are diffuse, tend to zero when B reduces to a singleton. The introduction of an appropriately coarse sub- σ -algebra on \mathbf{X} , such as the one generated by a given countable partition, would allow for recovering some of the results that hold for countable spaces. However, such an approach will not be natural or intuitive in most of the situations. Alternatively, if p and p' are absolutely continuous with respect to a reference measure $\lambda \in \mathbf{M}(\mathbf{X})$, then the probability density \hat{f} of $\hat{p}(\cdot | \Delta)$ can be expressed as a function of the probability densities f and f' of p and p' respectively as

$$\hat{f} : x \mapsto \frac{f(x)f'(x)}{\int f(y)f'(y)\lambda(dy)}.$$

However, to resort to this approach amounts to ignoring the incompatibility between the two random variables, in addition to the loss of meaningful interpretation of the denominator which becomes dependent on the choice of the reference measure. In a filtering context, one of the laws, say p , would be the prior whereas p' would represent the observation. Although the prior might be appropriately represented by a probability measure, the uncertainty in the observation is often mostly due to lack of knowledge for which a probability measure might be ill-suited.

Example 5. Attributing a probability measure to data originated from natural language is often inappropriate. For instance, if an observer locates an object in the real world *around* position $x \in \mathbb{R}^3$, then representing this data with a distribution on \mathbb{R}^3 centred on x , such as a Gaussian distribution with a given variance, highly overstates the given information. The observer is not giving the exact probability for the object to be in any measurable subset of \mathbb{R}^3 ; he only gives the fact that the probability should be low for subsets that are far from x in a given sense and high when close to or when containing x .

Overall, there is a need for the introduction of additional concepts that could account for these non-informative types of knowledge, and which would in turn enable data assimilation to be performed in more general spaces. The objective in the next section is to find an alternative way of representing uncertainty while staying in the standard formalism of measure and probability theory.

II. FUNDAMENTAL CONCEPTS

We first recall two concepts of measure theory that will be useful in this article, namely the concepts of outer measure and of convolution of measures.

A. Outer measure

The concept of *outer measure* is fundamental in measure theory and is defined as follows.

Definition 1. An outer measure on \mathbf{E} is a function $\mu : \wp(\mathbf{E}) \rightarrow [0, \infty]$ verifying the following conditions:

- a) (Null empty set) $\mu(\emptyset) = 0$
- b) (Monotonicity) if $A \subseteq B \subseteq \mathbf{E}$ then $\mu(A) \leq \mu(B)$
- c) (Countable sub-additivity) for every sequence $(A_n)_{n \in \mathbb{N}}$ of subsets of \mathbf{E}

$$\mu\left(\bigcup_{n \in \mathbb{N}} A_n\right) \leq \sum_{n \in \mathbb{N}} \mu(A_n).$$

Outer measures allow for constructing both σ -algebras and measures on them via Carathéodory's method [17, Sect. 113]: the outer measure μ induces a σ -algebra \mathcal{X} of subsets of \mathbf{X} composed of sets A verifying

$$(\forall C \subseteq \mathbf{X}) \quad \mu(C) = \mu(C \cap A) + \mu(C \cap A^c),$$

which are referred to as μ -measurable sets. The measure space $(\mathbf{X}, \mathcal{X}, \mu)$ is a complete measure space. A classical example of measures constructed in this way is the Lebesgue measure on the σ -algebra of Lebesgue measurable subsets of \mathbb{R} . If \mathbf{F} is a set, μ and μ' are outer measures on \mathbf{E} and \mathbf{F} respectively and $f : \mathbf{E} \rightarrow \mathbf{F}$ is a function, then

$$C \subseteq \mathbf{F} \mapsto \mu(f^{-1}[C]) \quad \text{and} \quad A \subseteq \mathbf{E} \mapsto \mu'(f[A])$$

are outer measures. One way of constructing outer measures from measures is as follows: let $(\mathbf{X}, \mathcal{X}, m)$ be a measure space, then m induces an outer measure m^* on \mathbf{X} defined as

$$m^*(C) = \inf\{\mu(A) \text{ s.t. } A \in \mathcal{X}, A \supseteq C\}, \quad (3)$$

for any $C \subseteq \mathbf{X}$.

B. Convolution of measures

An operation that will prove to be of importance is the operation of convolution of measures. We consider that the set \mathbf{X} is a Polish space equipped with its Borel σ -algebra $\mathcal{B}(\mathbf{X})$.

Definition 2. Let m and m' be two finite measures on a topological semigroup (\mathbf{X}, \cdot) , then the convolution $m * m'$ of m and m' is defined as

$$m * m'(A) = m \times m'(\{(x, y) \text{ s.t. } x \cdot y \in A\}), \quad (4)$$

for any $A \in \mathcal{B}(\mathbf{X})$.

The set function $m * m'$ defined in (4) is a measure on $(\mathbf{X}, \mathcal{B}(\mathbf{X}))$. The following properties of the operation of convolution are corollaries of [17, Sect. 444A-D].

Corollary 1. If (\mathbf{X}, \cdot) is a topological semigroup, then it holds that

$$m * (m' * m'') = (m * m') * m''$$

for all finite measures m , m' and m'' on \mathbf{X} . If (\mathbf{X}, \cdot) is commutative, then it holds that $m * m' = m' * m$ for all finite measures m and m' on \mathbf{X} .

III. MEASURE CONSTRAINT

Henceforth, we consider a space \mathbf{X} which is assumed to be a Polish space. We denote $\mathbf{L}^0(\mathbf{X}, \mathbb{R})$ the set of all measurable functions from $(\mathbf{X}, \mathcal{B}(\mathbf{X}))$ to $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ and we use the shorthand notation $\mathbf{L}^0(\mathbf{X}) \doteq \mathbf{L}^0(\mathbf{X}, \mathbb{R}^+)$ for the subset of $\mathbf{L}^0(\mathbf{X}, \mathbb{R})$ made of non-negative functions. The general definition of measure constraint is given below, followed by the introduction of more specific properties and operations.

A. Definition of measure constraint

Using the notion of outer measure defined in the previous section as well as the technical results about the set $\mathbf{L}^0(\mathbf{X}, \mathbb{R})$ detailed in [17, Sect. 245], we introduce the concept of *measure constraint* as follows.

Definition 3. Let M be a measure on $\mathbf{L}^0(\mathbf{X}, \mathbb{R})$, if it holds that the function μ_M defined on the power set $\wp(\mathbf{X})$ of \mathbf{X} as

$$\mu_M : A \mapsto M(\chi(A, \cdot)) \quad (5)$$

is an outer measure for a given collection of measurable functions $\{\chi(A, \cdot)\}_{A \subseteq \mathbf{X}}$ on $\mathbf{L}^0(\mathbf{X}, \mathbb{R})$, then M is said to be a *measure constraint* on \mathbf{X} with *characteristic function* χ . If m is a finite measure on \mathbf{X} verifying

$$m(B) \leq M(\chi(B, \cdot)) \quad (6)$$

for any $B \in \mathcal{B}(\mathbf{X})$ and

$$m(\mathbf{X}) = M(\chi(\mathbf{X}, \cdot)), \quad (7)$$

then M is said to be dominating m .

The motivation behind the introduction of measure constraints is to partially describe a measure by limiting the mass in some areas while possibly leaving it unconstrained elsewhere. Measure constraints that would bound a measure from below could also be defined using the associated concept of *inner measure*. In general, a measure on \mathbf{X} could be dominated by a measure on the set $\mathbf{L}^0(\mathbf{Y}, \mathbb{R})$ of measurable functions on a different set \mathbf{Y} , as long as the associated characteristic function χ is defined accordingly.

Remark 1. In Definition 3, the condition that μ_M is an outer measure is used to reduce the set of measures on $\mathbf{L}^0(\mathbf{X}, \mathbb{R})$ that would verify (6) to the ones that have natural properties. As explained by [17, Sect. 113B]: “The idea of the *outer measure* of a set A is that it should be some kind of upper bound for the possible measure of A ”. In fact, the use of outer measures as a way of dealing with uncertainty has first been proposed by [15]. In particular, the condition of monotonicity imposes that if a given mass is allowed in a set A then at least the same mass should be allowed in a larger set $B \supseteq A$. Similarly, the condition of sub-additivity allows for reaching the maximum mass $m(\mathbf{X})$ in several disjoint sets while still verifying $M(\chi(\mathbf{X}, \cdot)) = m(\mathbf{X})$.

Remark 2. A direct consequence of (6) is that a lower bound for $m(B)$ is also available for any $B \in \mathcal{B}(\mathbf{X})$ and is found to be

$$m(B) = m(\mathbf{X}) - m(B^c) \geq m(\mathbf{X}) - M(\chi(B^c, \cdot)). \quad (8)$$

The information provided by this lower bound is limited since μ_M is sub-additive and might reach $m(\mathbf{X})$ on any given set $B' \subset \mathbf{X}$. In this case, (8) only implies that $m(B') \geq 0$ which is not informative.

We will be particularly interested in the situation where m is a probability measure, in which case M satisfies $M(\chi(\mathbf{X}, \cdot)) = m(\mathbf{X}) = 1$ and is said to be a *probabilistic constraint*. The advantage with condition (7) is that if m is a finite measure that is not a probability measure, then m and M can be renormalised to be respectively a probability measure and a probabilistic constraint by dividing the inequality (6) by the total mass $m(\mathbf{X})$. When uncertainty is induced by a lack of knowledge on the actual law of a given random experiment then this law should be dominated by the considered probabilistic constraint. However, in cases of uncertain but non-random experiments, there is no such thing as a “true” law and the probabilistic constraint becomes the primary representation of uncertainty.

A useful case is found when M is supported by the set $\mathbf{L}^\infty(\mathbf{X})$ of non-negative bounded measurable functions and, assuming that $\sup : \mathbf{L}^\infty(\mathbf{X}) \rightarrow \mathbb{R}$ is measurable, when χ is such that

$$\chi : (A, f) \mapsto \|\mathbf{1}_A \cdot f\|, \quad (9)$$

where $\mathbf{1}_A$ denotes the indicator of A and where $\|\cdot\|$ is the uniform norm on $\mathbf{L}^\infty(\mathbf{X})$. All outer measures do not take the form assumed in (5) with χ as in (9), but this case offers suitably varied configurations by combining a linear part and a very sub-additive part, that is the measure by M and the uniform norm respectively. This case will be understood as the default situation in the sense that (9) will be considered when the characteristic function of a measure constraint is not specified. The subset of probabilistic constraints with such a supremum-based characteristic function is denoted $\mathbf{C}_1(\mathbf{X})$ and we consider the weighted semi-norm $\|\cdot\|$ defined on this set as

$$\|\cdot\| : M \mapsto \int \|f\| M(df).$$

This semi-norm is not a norm since $\|M\| = 0$ does not imply that M is the null measure in general.

B. Properties of measure constraints

Henceforth, $\mathcal{L}^\infty(\mathbf{X})$ will denote the Borel σ -algebra induced on $\mathbf{L}^\infty(\mathbf{X})$ by the topology of convergence in measure on $\mathbf{L}^0(\mathbf{X})$ presented in [17, Sect. 245]. The characterisation of $\mathcal{L}^\infty(\mathbf{X})$ is highly technical and out of the scope of this article so that the measurability of the function $f \mapsto \sup f$ and the mapping $f \mapsto f \circ \xi$, for some measurable mapping ξ in \mathbf{X} , is assumed rather than demonstrated.

Given the definition of $\mathbf{C}_1(\mathbf{X})$, it appears that many different measures in this set will dominate exactly the same probability measures in $\mathbf{M}_1(\mathbf{X})$ since a rescaling of the functions in the support can be compensated for by a rescaling of the measure itself. However such a multiplicity can be avoided by identifying a canonical form and a rescaling procedure as in the following proposition.

Proposition 1. For any probabilistic constraint $M \in \mathbf{C}_1(\mathbf{X})$, there exists a probabilistic constraint $M^\dagger \in \mathbf{C}_1(\mathbf{X})$ that is equivalent to M and is also a probability measure, which can be determined via the following rescaling procedure:

$$(\forall F \in \mathcal{L}^\infty(\mathbf{X})) \quad M^\dagger(F) \doteq \int \mathbf{1}_F(f^\dagger) \|f\| M(df), \quad (10)$$

where the function $f^\dagger \in \mathbf{L}^\infty(\mathbf{X})$ is defined as

$$f^\dagger \doteq \begin{cases} \frac{f}{\|f\|} & \text{if } \|f\| \neq 0 \\ 1 & \text{otherwise.} \end{cases}$$

Proof. By construction, it holds that the support of M^\dagger is included in the subset of $\mathbf{L}^\infty(\mathbf{X})$ made of functions with uniform norm equal to one, so that $M^\dagger(\mathbf{L}^\infty(\mathbf{X})) = \|M^\dagger\| = 1$. The measure M^\dagger is then both a probability measure and a probabilistic constraint. We now have to show that M and M^\dagger dominate the same probability measures: Let $p \in \mathbf{M}_1(\mathbf{X})$ be a probability measure dominated by M , then

$$m(B) \leq \int \|\mathbf{1}_B \cdot f\| M(df) = \int \|\mathbf{1}_B \cdot f^\dagger\| M'(df),$$

for all $B \in \mathcal{B}(\mathbf{X})$, where $M'(df) \doteq \|f\| M(df)$ so that, by a change of variable,

$$m(B) \leq \int \|\mathbf{1}_B \cdot f^\dagger\| M'(df) = \int \|\mathbf{1}_B \cdot f'\| M^\dagger(df'),$$

which terminates the proof. \square

Remark 3. The definition of f^\dagger when $\|f\| = 0$ is irrelevant because of the form of (10). Yet, considering $f^\dagger = 1$ when $\|f\| = 0$ implies that f^\dagger is always in the subset

$$\mathbf{L}(\mathbf{X}) \doteq \{f \in \mathbf{L}^\infty(\mathbf{X}) \text{ s.t. } \|f\| = 1\}$$

of measurable functions with uniform norm equal to one. Probabilistic constraints in the set $\mathbf{C}_1^*(\mathbf{X}) \doteq \mathbf{M}_1(\mathbf{L}(\mathbf{X}))$ will be referred to as *canonical probabilistic constraints*.

The unary operation of Proposition 1 does not affect the norm, i.e. the equality $\|M^\dagger\| = \|M\|$ holds by construction for any M in $\mathbf{C}_1(\mathbf{X})$. It is also *idempotent* and it distributes over the product \rtimes , that is $(M^\dagger)^\dagger = M^\dagger$ and $(M \rtimes M')^\dagger = M^\dagger \rtimes M'^\dagger$ hold for any $M, M' \in \mathbf{C}_1(\mathbf{X})$. Moreover, any canonical probabilistic constraint $P \in \mathbf{C}_1^*(\mathbf{X})$ verifies $P^\dagger = P$.

Remark 4. If $M \in \mathbf{C}_1(\mathbf{X})$ is a probabilistic constraint and if M^\dagger is supported by the set $\mathbf{I}(\mathbf{X})$ of measurable indicator functions, then the conditions (6) and (7) can be replaced by “ m agrees with the measure induced by μ_M ”, i.e. $m(B) = \mu_M(B)$ for any B in the σ -algebra of μ_M -measurable subsets.

Proposition 2. If $X, X' \in \mathbf{L}^0(\Omega, \mathbf{X})$ are two independent random variables on \mathbf{X} with respective laws p and p' and if M and M' are probabilistic constraints for p and p' respectively, then the joint law $p \rtimes p' \in \mathbf{M}_1(\mathbf{X})$ verifies

$$p \rtimes p'(\hat{B}) \leq \int \|\mathbf{1}_{\hat{B}} \cdot (f \rtimes f')\| M(df) M'(df')$$

for any $\hat{B} \in \mathcal{B}(\mathbf{X}) \otimes \mathcal{B}(\mathbf{X})$.

Proposition 2 is only an implication since a joint probability measure dominated by a probabilistic constraint of the same form might not correspond to independent random variables. This means that another concept is needed to describe this special class of joint probabilistic constraints.

Definition 4. Let $X \in \mathbf{L}^0(\Omega, \mathbf{X})$ and $X' \in \mathbf{L}^0(\Omega', \mathbf{X})$ be two random variables then X and X' are said to be *independently constrained* by $\hat{M} \in \mathbf{C}_1(\mathbf{X} \times \mathbf{X})$ if there exist $M, M' \in \mathbf{C}_1(\mathbf{X})$ such that \hat{M} can be expressed as a convolution on the semigroup $(\mathbf{L}^\infty(\mathbf{X}), \rtimes)$ as

$$\hat{M} = M * M', \quad (11)$$

recalling that, in this case,

$$M * M'(F) \doteq \int \mathbf{1}_F(f \rtimes f') M(df) M'(df')$$

for any measurable subset F of $\mathbf{L}^\infty(\mathbf{X} \times \mathbf{X})$.

Definition 4 implies that any function \hat{f} in the support of \hat{M} is such that there exist f and f' in $\mathbf{L}^\infty(\mathbf{X})$ for which $\hat{f} = f \rtimes f'$. Functions of this type can be said to be *separable*. The concept of independently constrained random variables introduced in Definition 4 differs in general from the standard concept of independence since, even if X and X' were defined on the same probability space, then *a)* independently constrained random variables might actually be correlated, and *b)* independent random variables that are not known to be independent might be represented by a probabilistic constraint that does not exclude correlation. However, the concepts coincide when the involved probabilistic constraints are equivalent to probability measures. In cases where the probability spaces on which these random variables are defined are different, independence or correlations cannot even be defined.

In order to illustrate the use of measure constraints for modelling uncertainty, several examples of probabilistic constraints are given in the next section.

IV. EXAMPLES OF PROBABILISTIC CONSTRAINTS

Throughout this section, $X \in \mathbf{L}^0(\Omega, \mathbf{X})$ will denote a random variable and $P \in \mathbf{C}_1^*(\mathbf{X})$ will be assumed to dominate its law $p \doteq X_* \mathbb{P}$.

A. Uninformative case

If $P = \delta_1$ then the only information provided by P on X is that

$$p(B) \leq \int \|\mathbf{1}_B \cdot f\| P(df) = 1$$

for any $B \in \mathcal{B}(\mathbf{X})$, which is non-informative since p is already known to be a probability measure. In other words, nothing is known about the random variable X . For instance, the σ -algebra of μ_P -measurable sets is the trivial σ -algebra $\{\emptyset, \mathbf{X}\}$.

B. Indicator function

If there exists $A \in \mathcal{B}(\mathbf{X})$ such that $P = \delta_{1_A}$, then it holds that

$$p(B) \leq \|\mathbf{1}_{A \cap B}\| = \begin{cases} 1 & \text{if } A \cap B \neq \emptyset \\ 0 & \text{otherwise,} \end{cases}$$

for any $B \in \mathcal{B}(\mathbf{X})$. As a probability measure, p is always less or equal to 1, so that the only informative part in the previous inequality is:

$$(\forall B \in \mathcal{B}(\mathbf{X})) \quad (A \cap B = \emptyset) \Rightarrow (p(B) = 0),$$

that is, all the probability mass of B lies within A . This means that the random variable X is only known to be in A almost surely. A uniform distribution over A , assuming it can be defined, would not model the same type of knowledge as the corresponding interpretation would be “realisations of X are equally likely everywhere inside A ”, whereas the interpretation associated with the probabilistic constraint δ_{1_A} is “whatever the law of X , it is only known that realisations will be inside A ”.

Remark 5. The σ -algebra of μ_P -measurable subsets is found to be the completion of the σ -algebra $\mathcal{X}_B \doteq \{\emptyset, B, B^c, \mathbf{X}\}$ with respect to any measure m on $(\mathbf{X}, \mathcal{X}_B)$ verifying $m(B) > 0$ as well as $m(B^c) = 0$.

C. Upper bound

If there exists $f \in \mathbf{L}(\mathbf{X})$ such that $P = \delta_f$ then for any $B \in \mathcal{B}(\mathbf{X})$,

$$p(B) \leq \int \|\mathbf{1}_B \cdot f'\| P(df') = \int \|\mathbf{1}_B \cdot f'\| \delta_f(df') = \|\mathbf{1}_B \cdot f\|.$$

In this case, δ_f can be seen as the simplest non-trivial form of probabilistic constraint. The two previous examples are special cases with $f = \mathbf{1}$ and $f = \mathbf{1}_A$.

Remark 6. Probabilistic constraints of this form are equivalent to *possibility distributions* [25, 12, 13, 14] and are also related to the notion of *membership function* of a *fuzzy set* [33]. However, the approach considered in this work does not rely on the notion of fuzzy set and the form considered here is only used as a simple example of probabilistic constraint.

D. Combination of upper bounds

If the probabilistic constraint P is of the form $P = \sum_{i=1}^N a_i \delta_{f_i}$ then

$$p(B) \leq \int \|\mathbf{1}_B \cdot f'\| P(df') = \sum_{i=1}^N a_i \|\mathbf{1}_B \cdot f_i\| = \sum_{i=1}^N a_i \sup_B f_i$$

for any $B \in \mathcal{B}(\mathbf{X})$. This combination of upper bounds is not equivalent to a single bound in general and allows for modelling more accurate information. For instance, if $P = 0.5\delta_{1_B} + 0.5\delta_{1_{B'}}$ with $B \cap B' = \emptyset$, then one half of the probability mass of p is in B and the other half is in B' .

E. Constraint based on a partition

If there exists a measurable countable partition π of \mathbf{X} and if the probability measure P is of the form

$$P = \sum_{B \in \pi} q(B) \delta_{1_B},$$

where q is a probability measure on the sub- σ -algebra generated by π , then

$$(\forall B \in \pi) \quad p(B) = q(B),$$

i.e. the information available about p is the one embedded into q .

Proof. From the definition of probabilistic constraint, we deduce that the inequality $p(B) \leq q(B)$ holds for any $B \in \pi$. Assume there exists $B \in \pi$ such that $p(B) < q(B)$. This assumption implies that

$$p(\mathbf{X}) = \sum_{B \in \pi} p(B) < \sum_{B \in \pi} q(B) = q(\mathbf{X}) = 1,$$

which is a contradiction since p is a probability measure. \square

Following the construction of an outer measure from a measure described in (3), we define the outer measure q^* induced by q on \mathbf{X} as follows

$$(\forall C \subseteq \mathbf{X}) \quad q^*(C) = \inf\{q(B) \text{ s.t. } B \in \sigma(\pi), B \supseteq C\}.$$

Since π is a partition of \mathbf{X} , it is easy to check that the outer measure μ_P verifying

$$\mu_P(C) = \sum_{B \in \pi : B \cap C \neq \emptyset} q(B)$$

for all subsets C of \mathbf{X} is equal to the outer measure q^* . This example shows that probabilistic constraints are versatile enough to model the same level of information as with a sub- σ -algebra generated by a partition.

F. Probability measure

Given that the singletons of \mathbf{X} are measurable, if the support of P is in the subset $\mathbf{I}_s(\mathbf{X}) \subset \mathbf{I}(\mathbf{X})$ of indicator functions on singletons, then there exists a measure q in $\mathbf{M}_1(\mathbf{X})$ such that, for any $B \in \mathcal{B}(\mathbf{X})$, it holds that

$$p(B) \leq \int_{\mathbf{I}_s(\mathbf{X})} \|\mathbf{1}_B \cdot f\| P(df) = \int \mathbf{1}_B(x) q(dx) = q(B),$$

from which we conclude that $p = q$. The proof of the equality is very similar to the proof for constraints based on a partition.

G. Plausibility

If P has the form

$$P = \sum_{A \in \mathcal{A}} g(A) \delta_{1_A}$$

for some set \mathcal{A} of measurable subsets and some function $g : \mathcal{A} \rightarrow \mathbb{R}^+$, then

$$p(B) \leq \int_{\mathbf{I}(\mathbf{X})} \|\mathbf{1}_B \cdot f\| P(df) = \sum_{A \in \mathcal{A} : A \cap B \neq \emptyset} g(A)$$

for any $B \in \mathcal{B}(\mathbf{X})$, and the probabilistic constraint P reduces to a *plausibility* as defined in the context of Dempster-Shafer theory [10, 28]. A non-technical overview of the concepts of this theory is given by [30, Sect. 4.4]. The two previous examples can be seen as special cases of plausibility where g is a measure on $\mathcal{A} = \pi$ or $\mathcal{A} = \mathbf{I}_s(\mathbf{X})$.

The multiple cases given in this section show that probabilistic constraints can model information with various degrees of precision. In this work, we will be mainly interested in the most informative and the most uninformative cases.

V. OPERATIONS ON MEASURE CONSTRAINTS

It is essential to be able to adapt a probabilistic constraint when the underlying probability measure is transformed, such as when considering a pushforward for the considered measure with respect to a given measurable function. In the following sections, \mathbf{X}_1 and \mathbf{X}_2 are assumed to be Polish spaces equipped with their Borel σ -algebra.

A. Pushforward

The operation of pushforwarding measures from one measurable space to another is central to measure theory and should therefore be translated in terms of measure constraints. In other words, if M dominates the measure m and ξ is a measurable function, then what is the measure constraint dominating $\xi_* m$?

The objective is to find sufficient conditions for the existence of a measure constraint M' verifying

$$M'(\chi(B, \cdot)) = M(\chi(\xi^{-1}[B], \cdot)) \quad (12)$$

for all $B \in \mathcal{B}(\mathbf{X}_2)$, where χ is the characteristic function based on the supremum norm defined in (9). It first appears from the properties of outer measures described in Section II-A that $A \mapsto M(\chi(\xi^{-1}[A], \cdot))$ is an outer measure, so that M' is possibly dominating $\xi_* m$. The existence of such a measure constraint is proved in the next proposition.

Proposition 3. *Let M be a measure constraint on \mathbf{X}_1 and let ξ be a measurable mapping from \mathbf{X}_1 to \mathbf{X}_2 , then the measure constraint M' on \mathbf{X}_2 defined as the pushforward of M by the mapping T_ξ from $\mathbf{L}^\infty(\mathbf{X}_1)$ to $\mathbf{L}^\infty(\mathbf{X}_2)$ characterised by*

$$(\forall f \in \mathbf{L}^\infty(\mathbf{X}_1)) \quad T_\xi(f) : y \mapsto \sup_{\xi^{-1}[\{y\}]} f$$

verifies (12).

Proof. The first step is to rewrite the uniform norm over \mathbf{X}_1 in a suitable way as

$$\begin{aligned} \|\mathbf{1}_{\xi^{-1}[B]} \cdot f\| &= \sup_{y \in \mathbf{X}_2} \left(\sup_{x \in \xi^{-1}[\{y\}]} (\mathbf{1}_{\xi^{-1}[B]}(x) f(x)) \right) \\ &= \sup_{y \in \mathbf{X}_2} \left(\mathbf{1}_B(y) \sup_{\xi^{-1}[\{y\}]} f \right) \end{aligned}$$

for any $f \in \mathbf{L}^\infty(\mathbf{X}_1)$ and any $B \in \mathcal{B}(\mathbf{X}_2)$, where the second line is explained by the fact that the function $\mathbf{1}_{\xi^{-1}[B]}$ is constant over $\xi^{-1}[\{y\}]$, and is equal to $\mathbf{1}_B(y)$ everywhere on this subset. By [6, Corollary 2.13], it holds that $T_\xi(f) \in \mathbf{L}^\infty(\mathbf{X}_2)$ since $\{\xi^{-1}[\{y\}] \times \{y\} : y \in B\}$ is a measurable subset of

$\mathbf{X}_1 \times \mathbf{X}_2$ for any $B \in \mathcal{B}(\mathbf{X}_2)$. The fact that T_ξ is measurable follows from the assumption that $\sup : \mathbf{L}^\infty(\mathbf{X}) \rightarrow \mathbb{R}$ is measurable. It then holds that

$$\int \|\mathbf{1}_B \cdot f'\| M'(df') = \int \|\mathbf{1}_{\xi^{-1}[B]} \cdot f\| M(df),$$

where M' is defined as the pushforward $(T_\xi)_* M$. \square

Remark 7. Proposition 3 can be straightforwardly generalised to characteristic functions of the form $\chi : (B, f) \mapsto \|(\chi' \circ \mathbf{1}_B) \cdot f\|$ for some suitable function $\chi' : \mathbb{R} \rightarrow \mathbb{R}$.

In the following example, the operation of marginalisation is translated to probabilistic constraints.

Example 6. Let \mathbf{X} be the Cartesian product of \mathbf{X}_1 and \mathbf{X}_2 and let p be a given probability measure on \mathbf{X} . The objective is to project a given probabilistic constraint $P \in \mathbf{C}_1(\mathbf{X})$ for p into a probabilistic constraint P' in $\mathbf{C}_1(\mathbf{X}_2)$ for the corresponding marginal $p' \in \mathbf{M}_1(\mathbf{X}_2)$ characterised by $p'(B) = p(\mathbf{X}_1 \times B)$. Marginalisation can be performed on P' by using Proposition 3 with the canonical projection map $\xi : \mathbf{X} \rightarrow \mathbf{X}_2$, from which we find that

$$(\forall B \in \mathcal{B}(\mathbf{X}_2)) \quad p(B) \leq \int \|\mathbf{1}_B \cdot f'\| (T_\xi)_* P(df'),$$

where the mapping $T_\xi : \mathbf{L}^\infty(\mathbf{X}) \rightarrow \mathbf{L}^\infty(\mathbf{X}_2)$ is found to be

$$T_\xi(f) : y \mapsto \sup_{\xi^{-1}[\{y\}]} f = \|f(\cdot, y)\|.$$

Another case, studied in the next example, is the pushforward of a probabilistic constraint defined as being equivalent to a probability measure on a sub- σ -algebra.

Example 7. Considering again the case detailed in Section IV where π is a measurable countable partition of \mathbf{X}_1 and where the probabilistic constraint P is of the form

$$P = \sum_{B \in \pi} p'(B) \delta_{\mathbf{1}_B},$$

where p' is a probability measure on the sub- σ -algebra generated by π , then for any measurable mapping $\xi : \mathbf{X}_1 \rightarrow \mathbf{X}_2$ such that $\sigma(\xi) \subseteq \sigma(\pi)$, it holds that

$$(\forall B \in \pi) \quad T_\xi(\mathbf{1}_B) : y \mapsto \sup_{\xi^{-1}[\{y\}]} \mathbf{1}_B = \mathbf{1}_{\xi(B)}(y),$$

that is, since $\xi[B]$ is a singleton,

$$T_\xi(\mathbf{1}_B) = \mathbf{1}_{\xi(B)} \in \mathbf{I}_s(\mathbf{X}_2),$$

so that $(T_\xi)_* P$ has its support in $\mathbf{I}_s(\mathbf{X}_2)$ and is therefore equivalent to a probability measure on \mathbf{X}_2 .

The concepts of pushforward and marginalisation for probabilistic constraints are important and will be useful in practice. We now consider the converse operation, referred to as pull-back and which will also contribute to the formulation of the main results in this work.

B. Pullback

The objective in this section is to understand in which situations it is possible to reverse the operation of pushforwarding measures and measure constraints. We first define what is understood as a pullback.

Definition 5. Let m be a measure on the space \mathbf{X}_2 and let ξ be a measurable mapping from \mathbf{X}_1 to \mathbf{X}_2 , then a *pullback* measure, denoted ξ^*m , is a measure in $\mathbf{M}_1(\mathbf{X}_1)$ satisfying

$$\xi_*(\xi^*m) = m.$$

Note that there are possibly many pullback measures for a given measure so that the concept of measure constraint will be useful in order to represent this multiplicity, as in the following proposition. As with the pushforward, we want to obtain the existence of a measure constraint M' with characteristic function χ for the pullback of a measure $m \in \mathbf{M}_1(\mathbf{X}_2)$ such that

$$M'(\chi(\xi^{-1}[B], \cdot)) = M(\chi(B, \cdot)) \quad (14)$$

holds for any $B \in \mathcal{B}(\mathbf{X}_2)$, where M is a measure constraint dominating m . It appears from the properties of outer measures described in Section II-A that $A \mapsto M(\chi(\xi[A], \cdot))$ is an outer measure, so that M' is possibly a measure constraint dominating the pullback measures of m .

Proposition 4. Let M be a measure constraint on \mathbf{X}_2 and let ξ be a measurable mapping from \mathbf{X}_1 to \mathbf{X}_2 , then the measure constraint M' on \mathbf{X}_1 defined as the pushforward of M by the mapping $T'_\xi : \mathbf{L}^\infty(\mathbf{X}_2) \rightarrow \mathbf{L}^\infty(\mathbf{X}_1)$ defined as

$$T'_\xi : f \mapsto f \circ \xi$$

verifies (14).

Proof. We rewrite the uniform norm of $\mathbf{1}_B \cdot f$ as

$$\|\mathbf{1}_B \cdot f\| = \|\mathbf{1}_{\xi^{-1}[B]} \cdot (f \circ \xi)\|$$

for any $f \in \mathbf{L}^\infty(\mathbf{X}_2)$ and any $B \in \mathcal{B}(\mathbf{X}_2)$. Since the composition of measurable mappings is measurable, we verify that the codomain of T'_ξ is $\mathbf{L}^\infty(\mathbf{X}_1)$. Since the mapping T'_ξ is assumed to be measurable, we write

$$\int \|\mathbf{1}_B \cdot f\| M(df) = \int \|\mathbf{1}_{\xi^{-1}[B]} \cdot f'\| (T'_\xi)_* M(df'),$$

which terminates the proof of the proposition. \square

Remark 8. Other transformations than T'_ξ could lead to a measure on $\mathbf{L}^\infty(\mathbf{X}_1)$ that verifies (14). However, these measures would not dominate all the pullback measures of m .

Example 8. If m is a known probability measure, then the corresponding canonical probabilistic constraint M would have its support in the set $\mathbf{I}_s(\mathbf{X}_2)$ of indicator functions of singletons in \mathbf{X}_2 . The induced probabilistic constraint M' on \mathbf{X}_1 would then be of the form

$$M' = \sum_{B \in \pi} m'(B) \delta_{\mathbf{1}_B},$$

where π is the partition generated by ξ on \mathbf{X}_2 and m' is the probability measure induced by m on the sub- σ -algebra generated by ξ . This example closes a loop started in Example 7

where a probabilistic constraint on a sub- σ -algebra generated by a partition was shown to induce a probability measure on \mathbf{X}_2 .

Now equipped with the concepts of pushforward and pullback for measure constraints and measures, we can address the question of data assimilation between different sources of information.

VI. DATA ASSIMILATION WITH PROBABILISTIC CONSTRAINTS

The objective in this section is to introduce a data-assimilation operation for independent sources of information about the same physical system.

A. State-space fusion

We first address the case where the state satisfies some desirable but stringent conditions before showing that a much larger class of state spaces can be considered under a weak assumption. We assume that there exists a *representative* set \mathcal{X} in which the state of the system of interest can be uniquely characterised and we denote ξ the projection map between \mathcal{X} and \mathbf{X} . Both \mathcal{X} and \mathbf{X} are assumed to be Polish spaces equipped with their Borel σ -algebras.

Two probabilistic constraints $P, P' \in \mathbf{C}_1(\mathcal{X})$ are considered and assumed to be *compatible*, that is, at most one of them corresponds to an unknown probability measure. These probabilistic constraints can be understood as either prior knowledge and observation or both as observations, which justifies the use of the generic term *data* to cover the two types of information. Although one or both of these probabilistic constraint do not represent a true law, it is useful to consider two probability measures respectively dominated by P and P' in order to unveil the mechanism behind data assimilation.

Let $X \in \mathbf{L}^0(\Omega, \mathbf{X})$ and $X' \in \mathbf{L}^0(\Omega', \mathbf{X})$ be two random variables in \mathbf{X} constrained by P and P' respectively and consider the predicate

“ X and X' represent the same physical system”. (#)

The random variables X and X' are defined on different probability spaces as they are assumed to originate from different representations of randomness. As a consequence, the predicate (#) cannot be expressed as $X = X'$. A random variable $X \times X'$ can be defined on the probability space $(\Omega \times \Omega', \Sigma \otimes \Sigma', \mathbb{P} \times \mathbb{P}')$ of joint outcomes/events and the law of $X \times X'$ is the probability measure $p \times p'$ where p and p' are the respective laws of X and X' . The predicate (#) cannot be expressed as an event in $\mathbf{X} \times \mathbf{X}$ in general, yet, it can be expressed as the event $\Delta = \{(x, x) \text{ s.t. } x \in \mathcal{X}\}$ in the product σ -algebra $\mathcal{B}(\mathcal{X}) \otimes \mathcal{B}(\mathcal{X})$ because of the representativity of the set \mathcal{X} .

The data-assimilation operation for independent sources of information can now be formalised as an operation between probabilistic constraints as shown in the following theorem.

Theorem 1. Let \mathcal{X} be a representative set and let P and P' be compatible probabilistic constraints on \mathcal{X} assumed to represent the same physical system, then the probabilistic

constraint resulting from data assimilation is characterised by the following convolution of measures on $(\mathbf{L}^\infty(\mathcal{X}), \cdot)$:

$$P \star P' = \frac{(P * P')^\dagger}{\|P * P'\|}, \quad (15)$$

whenever $\|P * P'\| \neq 0$ holds.

The use of the unary operation \cdot^\dagger in the numerator of (15) has for only objective to make $P \star P'$ a canonical probabilistic constraint. An equivalent choice would be to define $P \star P'$ as the normalised version of $P * P'$. However, (15) is preferred since it ensures that $P \star P'$ is also a probability measure on $\mathbf{L}^\infty(\mathcal{X})$. A similar rule has been proposed without proof in [23] in the context of fuzzy Dempster-Shafer theory [32].

Proof. Let $X \in \mathbf{L}^0(\Omega, \mathcal{X})$ and $X' \in \mathbf{L}^0(\Omega', \mathcal{X})$ be two random variables in \mathcal{X} which laws $p = X_*\mathbb{P}$ and $p' = X'_*\mathbb{P}'$ are respectively dominated by P and P' . The main idea of the proof is to build the joint law $p \rtimes p'$ in $\mathcal{X} \times \mathcal{X}$ and to study the diagonal $\Delta = \{(x, x) \text{ s.t. } x \in \mathcal{X}\}$. The data assimilation mechanism for probabilistic constraints can then be deduced from the one for p and p' . Let $\Delta(B)$ be the intersection $\Delta \cap B \times B$ for any $B \in \mathcal{B}(\mathcal{X})$ and consider the law $p \rtimes p'$ verifying

$$p \rtimes p'(B \times B') \leq \int \|\mathbf{1}_{B \times B'} \cdot (f \rtimes f')\| P(df) P'(df'),$$

for any $B \times B' \in \mathcal{B}(\mathcal{X} \times \mathcal{X})$ so that

$$\begin{aligned} p \rtimes p'(\Delta(B)) &\leq \int \|\mathbf{1}_{\Delta(B)} \cdot (f \rtimes f')\| P(df) P'(df') \\ &\leq \int \|\mathbf{1}_B \cdot (f \cdot f')^\dagger\| \|f \cdot f'\| P(df) P'(df') \\ &\leq \int \|\mathbf{1}_B \cdot \hat{f}\| M(d\hat{f}) \end{aligned}$$

holds for any $B \in \mathcal{B}(\mathcal{X})$, with

$$M(F) \doteq \int \mathbf{1}_F((f \cdot f')^\dagger) \|f \cdot f'\| P(df) P'(df') = (P * P')^\dagger(F)$$

for any measurable subset F of $\mathbf{L}^\infty(\mathcal{X})$. If both p and p' were the actual probability laws depicting the two sources of information, it would be necessary to find a lower bound for $p \rtimes p'(\Delta)$ in order to determine the constraint for the hypothetical posterior law given Δ (note that such a lower bound would be equal to zero in most of the situations). On the contrary, it has been assumed that at least one of the probabilistic constraints is a primary representation of uncertainty so that it is the measure constraint M that has to be normalised. We conclude that the probabilistic constraint $P \star P'$ on \mathcal{X} verifies

$$P \star P'(F) = \frac{1}{C} (P * P')^\dagger(F)$$

for any $F \in \mathcal{L}^\infty(\mathcal{X})$, where the normalising constant is found to be

$$C = (P * P')^\dagger(\mathbf{L}^\infty(\mathbf{X})) = \int \|f \cdot f'\| P(df) P'(df') = \|P * P'\|.$$

The convolution of P and P' is well defined since $(\mathbf{L}^\infty(\mathcal{X}), \cdot)$ is a topological semigroup, as shown in [17, Sect. 245D]. \square

Remark 9. Using the notations of Theorem 1 and defining the likelihood function $\ell(\Delta | \cdot)$ on $\mathbf{L}^\infty(\mathcal{X}) \times \mathbf{L}^\infty(\mathcal{X})$ as

$$\ell(\Delta | \cdot) : (f, f') \mapsto \|f \cdot f'\|,$$

the probability measure $P \star P'$ can be expressed for any $F \in \mathcal{L}^\infty(\mathcal{X})$ as

$$P \star P'(F) = \frac{P \rtimes P'(\ell(\Delta | \cdot) \Phi(\cdot, F))}{P \rtimes P'(\ell(\Delta | \cdot))},$$

where Φ is a Markov kernel from $\mathbf{L}^\infty(\mathcal{X}) \times \mathbf{L}^\infty(\mathcal{X})$ to $\mathbf{L}^\infty(\mathcal{X})$ defined as

$$\Phi : ((f, f'), F) \mapsto \delta_{(f \cdot f')^\dagger}(F).$$

This shows that the canonical version of $P \star P'$ is a proper Bayes' posterior probability measure on $\mathbf{L}^\infty(\mathcal{X})$. This is a fundamental result since it shows that general data assimilation can be performed in a fully measure-theoretic Bayesian paradigm.

One immediate objection to Theorem 1 is that the assumption that the set \mathcal{X} is representative is highly restrictive. Very often there is no interest in handling sophisticated state spaces, especially when the received information does not allow for such an in-depth representation. However, several systems could have the same state in \mathbf{X} and $(\#)$ cannot be expressed as an event in $\mathcal{B}(\mathbf{X}) \otimes \mathcal{B}(\mathbf{X})$ so that a different approach has to be considered. In the next corollary, we show that the result of Theorem 1 can be extended to a simpler state space as long as it verifies a natural assumption.

Corollary 2. *Assuming that there exists a representative set in which the physical system can be uniquely characterised, the result of Theorem 1 holds for random variables in the state space $(\mathbf{X}, \mathcal{B}(\mathbf{X}))$.*

Proof. Let ξ be a given projection map from \mathcal{X} to \mathbf{X} . The operations of pushforward and pullback can be respectively expressed via the following mappings on the space of measurable functions:

$$(\forall f \in \mathbf{L}^\infty(\mathcal{X})) \quad T_\xi(f) : x \mapsto \sup_{\xi^{-1}[\{x\}]} f,$$

and

$$(\forall f \in \mathbf{L}^\infty(\mathbf{X})) \quad T'_\xi(f) = f \circ \xi.$$

We compute the result of the pushforward of combined pullback functions as follows:

$$(\forall f, f' \in \mathbf{L}^\infty(\mathbf{X})) \quad T_\xi(T'_\xi(f) \cdot T'_\xi(f')) = T_\xi((f \cdot f') \circ \xi),$$

so that

$$T_\xi(T'_\xi(f) \cdot T'_\xi(f')) : x \mapsto \sup_{\xi^{-1}[\{x\}]} T_\xi((f \cdot f') \circ \xi) = f \cdot f'.$$

This result indicates that the pushforward and the pullback via ξ do not need to be considered when combining probabilistic constraints on the non-representative state space \mathbf{X} and that, as a result, data assimilation with probabilistic constraints on \mathbf{X} and on \mathcal{X} can be performed in the same way. \square

The following examples make use of the notations of Theorem 1 and detail two simple cases for which only one

of the elements in the probabilistic constraint is maintained through the data assimilation operation.

Example 9. If P has its support in $\mathbf{I}_s(\mathbf{X})$ and is therefore equivalent to $p \in \mathbf{M}_1(\mathbf{X})$ and if P' is of the form $P' = \delta_{f'}$, then the probabilistic constraint $P \star P' \in \mathbf{C}_1(\mathbf{X})$ satisfies

$$P \star P'(F) \propto \int \mathbf{1}_F((f'(x)\mathbf{1}_{\{x\}})^\dagger) f'(x) p(dx).$$

Since $f'(x)\mathbf{1}_{\{x\}}$ is either the null function or is supported by a singleton, $P \star P'$ has its support in $\mathbf{I}_s(\mathbf{X})$ and is equivalent to the measure $\hat{p} \in \mathbf{M}_1(\mathbf{X})$ defined as

$$\hat{p}(dx) \stackrel{f}{=} \frac{1}{p(f')} f'(x) p(dx),$$

that is, the law of the combined random variable is known to be \hat{p} . The same type of result holds if the roles of P and P' are interchanged. If f' is actually a likelihood of the form ℓ_z , then \hat{p} can be expressed as

$$\hat{p}(dx) \stackrel{f}{=} \frac{\ell_z(x) p(dx)}{\int \ell_z(x') p(dx')},$$

which is the usual Bayes' posterior of p given z , where z is interpreted as an observation and p is the prior probability measure. The fact that ℓ_z is an element of $\mathbf{L}(\mathbf{X})$ rather than a probability density does not affect the result because of the normalisation factor $\int \ell_z(x') p(dx')$. As mentioned in Example 3, one convenient choice for ℓ_z is the Gaussian-shaped upper bound defined in the one-dimensional observation case as

$$\ell_z(x) \doteq \exp\left(-\frac{(Hx - z)^2}{2\sigma^2}\right),$$

where σ corresponds to the standard deviation of the corresponding normal distribution and where H is the observation matrix. This choice allows for keeping the Kalman filter recursion while better representing the information received via finite-resolution sensors. Solutions to the multi-target tracking problem in the linear Gaussian case can be formulated in terms of Kalman filters in interaction [7, 8] and depend on the value of the denominator in Bayes' theorem which would usually be of the form

$$\frac{1}{\sqrt{2\pi}\varsigma} \exp\left(-\frac{(Hm - z)^2}{2\varsigma^2}\right),$$

with $\varsigma^2 = HPH^T + \sigma^2$, where m and P are the mean of variance of the prior distribution and where \cdot^T is the matrix transposition. With the proposed solution, we find that

$$\int \ell_z(x) p(dx) = \frac{\sigma}{\varsigma} \exp\left(-\frac{(Hm - z)^2}{2\varsigma^2}\right),$$

which is dimensionless, takes value in the interval $[0, 1]$ and can be interpreted as the probability for the observation z to belong to the target with distribution p .

Example 10. If P and P' have the form $P = \delta_f$ and $P' = \delta_{f'}$, then the probabilistic constraint $P \star P' \in \mathbf{C}_1(\mathbf{X})$ is found to be

$$P \star P' = \delta_{(f \cdot f')^\dagger}$$

that is, $P \star P'$ is a canonical probabilistic constraint and $(f \cdot f')^\dagger \in \mathbf{L}(\mathbf{X})$ can be seen as the resulting upper bound.

The binary operation \star on $\mathbf{C}_1(\mathbf{X})$ introduced in Theorem 1 has some additional properties that are detailed in the following theorem.

Theorem 2. *The space $(\mathbf{C}_1(\mathbf{X}), \star)$ is a commutative semi-group with an identity element.*

Proof. Theorem 1 together with Corollary 2 show that \star is a proper binary operation on $\mathbf{C}_1(\mathbf{X})$ in the sense that it is found to be a relation from $\mathbf{C}_1(\mathbf{X}) \times \mathbf{C}_1(\mathbf{X})$ to $\mathbf{C}_1(\mathbf{X})$. In order to prove the result of the theorem, we need to show that \star is also associative, has an identity element in $\mathbf{C}_1(\mathbf{X})$ and is commutative: let P, P' and P'' be probabilistic constraints in $\mathbf{C}_1(\mathbf{X})$, then

- a) \star is associative: Since the convolution of measures $*$ is known to be associative, we need to show that $((P \star P')^\dagger \star P'')^\dagger = (P \star (P' \star P''))^\dagger$ holds. This can be done by verifying that

$$\|f \cdot f'\| \|\hat{f} \cdot f''\| = \|f \cdot f' \cdot f''\|$$

and

$$\frac{\hat{f} \cdot f''}{\|\hat{f} \cdot f''\|} = \frac{f \cdot f' \cdot f''}{\|f \cdot f' \cdot f''\|}$$

hold with $\hat{f} = \frac{f \cdot f'}{\|f \cdot f'\|}$ for any $f, f', f'' \in \mathbf{L}^\infty(\mathbf{X})$.

- b) \star has an identity element: consider that $P' = \delta_1$, then for any measurable subset $F \in \mathcal{L}^\infty(\mathbf{X})$

$$P \star P'(F) \propto \int \mathbf{1}_F((f \cdot \mathbf{1})^\dagger) \|f\| P(df) = P(F),$$

with the coefficient of proportionality

$$\int \|f \cdot \mathbf{1}\| P(df) = 1,$$

so that $P \star P' = P$ and P' is the identity element of $(\mathbf{C}_1(\mathbf{X}), \star)$.

This terminates the proof. \square

The absence of inverse element for the operator \star on $\mathbf{C}_1(\mathbf{X})$ is due to the fact that it is not always possible to “forget” what has been learnt. We now study the relation between the proposed operation and Dempster's rule of combination in the next example.

Example 11. Let P and P' be two canonical probabilistic constraints on \mathbf{X} of the form

$$P = \sum_{A \in \mathcal{A}} m(A) \delta_{\mathbf{1}_A} \quad \text{and} \quad P' = \sum_{A' \in \mathcal{A}'} m'(A') \delta_{\mathbf{1}_{A'}}$$

for some sets \mathcal{A} and \mathcal{A}' of measurable subsets and some functions $m : \mathcal{A} \rightarrow [0, 1]$ and $m' : \mathcal{A}' \rightarrow [0, 1]$. The posterior probabilistic constraint $P \star P'$ on \mathbf{X} verifies

$$P \star P'(F) \propto \sum_{A \cap A' \neq \emptyset} \mathbf{1}_F(\mathbf{1}_{A \cap A'}) m(A) m'(A')$$

for any $F \in \mathcal{L}^\infty(\mathbf{X})$, which can be re-expressed as

$$P \star P' = \frac{1}{\|P \star P'\|} \sum_{A \cap A' \neq \emptyset} m(A) m'(A') \delta_{\mathbf{1}_{A \cap A'}},$$

with

$$\|P * P'\| = \sum_{A \cap A' \neq \emptyset} m(A)m'(A') = 1 - \sum_{A \cap A' = \emptyset} m(A)m'(A'),$$

so that $P \star P'$ reduces to the result of Dempster's rule of combination between m and m' [11, 29]. A discussion about the connections between Dempster's rule of combination and Bayes' theorem applied to second-order probabilities can be found in [1].

B. General fusion

In Section VI-A, it has been assumed that the set \mathbf{X} on which the probabilistic constraints are defined can be understood as a state space, with the consequence that the system of interest should be represented at a single point of \mathbf{X} . The way elements of $\mathbf{L}^\infty(\mathbf{X})$ are fused is highly dependent on this assumption. Yet, a different viewpoint must be considered when introducing probabilistic constraints on more general sets such as $\mathbf{M}_1(\mathbf{X})$ and $\mathbf{C}_1(\mathbf{X})$.

Remark 10. In order to have a suitable topology defined on $\mathbf{L}^0(\mathbf{C}_1(\mathbf{X}))$, a reference measure must be introduced. As there is no natural reference measure on $\mathbf{C}_1(\mathbf{X})$, a case-by-case reference measure must be defined, e.g., as an integer-valued measure that singles out a countable number of elements of $\mathbf{C}_1(\mathbf{X})$, or as a parametric family of elements of $\mathbf{C}_1(\mathbf{X})$ which parameter lies in an Euclidean space.

Let \mathbf{Y} be a topological space equipped with its Borel σ -algebra and assume there is a given way of combining elements of \mathbf{Y} which is characterised by a potential function $\ell : \mathbf{Y} \times \mathbf{Y} \rightarrow [0, 1]$ and a surjective mapping $\theta : S \rightarrow \mathbf{Y}$, with $S \subseteq \mathbf{Y} \times \mathbf{Y}$, which defines a stochastic kernel Φ from $\mathbf{Y} \times \mathbf{Y}$ to \mathbf{Y} as

$$\Phi((y, y'), \cdot) \doteq \begin{cases} \delta_{\theta(y, y')} & \text{if } (y, y') \in S \\ \mathbf{0} & \text{otherwise.} \end{cases}$$

We also assume that ℓ verifies $\ell(y, y') = 0$ for any $(y, y') \in S^c$. Note that the kernel Φ becomes a Markov kernel when restricted to S . When it exists, the probability measure \hat{p} in $\mathbf{M}_1(\mathbf{Y})$ defined as

$$\hat{p}(B) = \frac{\mathbf{p} \rtimes \mathbf{p}'(\ell\Phi(\cdot, B))}{\mathbf{p} \rtimes \mathbf{p}'(\ell)}, \quad (17)$$

for all $B \in \mathcal{B}(\mathbf{Y})$ can be introduced. As in Remark 9, \hat{p} can be seen as the projection of the Bayes' posterior corresponding to the prior $\mathbf{p} \rtimes \mathbf{p}'$ and the likelihood ℓ .

Example 12. If \mathbf{Y} is equal to $\mathbf{L}^\infty(\mathcal{X})$, with \mathcal{X} the representative state space defined in Section VI-A, then the laws \mathbf{p} and \mathbf{p}' are probabilistic constraints on \mathcal{X} . In this case, the natural choice for ℓ and θ is $\ell(f, f') = \|f \cdot f'\|$ and $\theta(f, f') = (f \cdot f')^\dagger$ for any $f, f' \in \mathbf{L}^\infty(\mathcal{X})$. The probability measure \hat{p} then takes the form

$$\hat{p}(F) \propto \int \mathbf{1}_F((f \cdot f')^\dagger) \|f \cdot f'\| \mathbf{p}(df) \mathbf{p}'(df'),$$

for all $F \in \mathcal{L}^\infty(\mathbf{Y})$, that is $\hat{p} = \mathbf{p} \star \mathbf{p}'$.

We assume that there is a reference measure on \mathbf{Y} which enables probability measures to be defined on $\mathbf{L}^\infty(\mathbf{Y})$. The

objective is to extend the fusion of probabilistic constraints to the case where \mathbf{Y} is an arbitrary set.

Theorem 3. Let \mathbf{P} and \mathbf{P}' be probabilistic constraints on \mathbf{Y} assumed to represent the same physical system, then the probabilistic constraint resulting from data assimilation is characterised by the following convolution of measures on $(\mathbf{L}^\infty(\mathcal{X}), \odot)$:

$$\mathbf{P} \star \mathbf{P}' = \frac{(\mathbf{P} * \mathbf{P}')^\dagger}{\|\mathbf{P} * \mathbf{P}'\|}, \quad (18)$$

where the binary operation \odot on $\mathbf{L}^\infty(\mathbf{Y})$ characterised by

$$f \odot f' : \hat{y} \mapsto \sup_{(y, y') \in \theta^{-1}\{\hat{y}\}} \ell(y, y') f(y) f'(y') \quad (19)$$

is assumed to be associative.

The level of generality of Theorem 3 is necessary for tackling estimation problems for systems that display hierarchical levels of uncertainty, such as in the modelling of stochastic populations [20].

Proof. Let \mathbf{p} and \mathbf{p}' be probability measures on \mathbf{Y} dominated by \mathbf{P} and \mathbf{P}' respectively. Equation (17) shows that the measure \mathbf{m} to be dominated is characterised by

$$(\forall B \in \mathcal{B}(\mathbf{Y})) \quad \mathbf{m}(B) = \mathbf{p} \rtimes \mathbf{p}'(\ell\Phi(\cdot, B)),$$

and we find that

$$\mathbf{m}(B) \leq \int \sup_{(y, y') : \theta(y, y') \in B} (\ell(y, y') f(y) f'(y')) \mathbf{P}(df) \mathbf{P}'(df')$$

for all $B \in \mathcal{B}(\mathbf{Y})$. In order to express the right hand side of the previous inequality as a probabilistic constraint, the argument of the supremum has to be normalised, and we find that

$$\begin{aligned} \mathbf{P} \star \mathbf{P}'(F) &= \frac{1}{C} \int \mathbf{1}_F((f \odot f')^\dagger) \|f \odot f'\| \mathbf{P}(df) \mathbf{P}'(df') \\ &\propto (\mathbf{P} * \mathbf{P}')^\dagger(F) \end{aligned}$$

for all $F \in \mathcal{L}^\infty(\mathbf{Y})$, where the operation \odot is defined as in the statement of the theorem. The constant C is found to be

$$C = \int \|f \odot f'\| \mathbf{P}(df) \mathbf{P}'(df') = \|\mathbf{P} * \mathbf{P}'\|,$$

thus proving the result of the theorem. \square

For the posterior probabilistic constraint of Theorem 3 to be well defined, the mappings θ and ℓ have to have sufficient properties for $(\mathbf{L}^\infty(\mathbf{Y}), \odot)$ to be a semigroup. Informally, the mapping θ can be loosely seen as a binary operation which has to be associative for \odot to be associative. More rigorously, we can define an extension $\tilde{\mathbf{Y}} \doteq \mathbf{Y} \cup \{\varphi\}$ of the set \mathbf{Y} by an isolated point φ and extend θ and ℓ as follows: $\theta(y, y') = \varphi$ for any $(y, y') \notin S$ and $\ell(y, y') = 0$ if $(y, y') \notin \mathbf{Y} \times \mathbf{Y}$. With these notations, we can formulate the following proposition.

Proposition 5. The binary operation \odot is associative if θ is associative and if

$$\ell(y, y') \ell(\theta(y, y'), y'') = \ell(y, \theta(y', y'')) \ell(y', y'')$$

holds for any $y, y', y'' \in \mathbf{Y}$.

Proof. We have to show that $(f \odot f') \odot f'' = f \odot (f' \odot f'')$ for any $f, f', f'' \in \mathbf{L}^\infty(\mathbf{Y})$ under the conditions given in the proposition. It holds that for any $\tilde{y} \in \mathbf{Y}$

$$\begin{aligned} ((f \odot f') \odot f'')(\tilde{y}) = \\ \sup_{\substack{(\hat{y}, y'') \in \theta^{-1}[\{\tilde{y}\}] \\ (y, y') \in \theta^{-1}[\{\hat{y}\}]} } \ell(y, y') \ell(\hat{y}, y'') f(y) f'(y') f''(y''), \end{aligned}$$

which can be expressed as

$$\begin{aligned} ((f \odot f') \odot f'')(\tilde{y}) = \\ \sup_{(y, y', y'') : \theta(\theta(y, y'), y'') = \tilde{y}} \ell(y, y') \ell(\theta(y, y'), y'') f(y) f'(y') f''(y''). \end{aligned}$$

Obtaining the equivalent expression for $f \odot (f' \odot f'')$ confirms that \odot is associative if

$$\begin{aligned} \theta(\theta(y, y'), y'') &= \theta(y, \theta(y', y'')) \\ \ell(y, y') \ell(\theta(y, y'), y'') &= \ell(y, \theta(y', y'')) \ell(y', y'') \end{aligned}$$

holds for any $y, y', y'' \in \mathbf{Y}$, the first line being equivalent to the associativity of θ when seen as a binary operation. \square

Notice that when the mapping θ is bijective, the function $f \odot f'$ is also characterised by the relation

$$f \odot f'(\theta(y, y')) = \ell(y, y') f(y) f'(y') \quad (22)$$

for all $(y, y') \in S$. Indeed, the supremum in (19) is taken over $\theta^{-1}[\{\tilde{y}\}]$ which is a singleton when θ is bijective. The following example makes the connection between state-space fusion and the more general fusion operation introduced in Theorem 3.

Example 13. If \mathbf{Y} is equal to the state space \mathbf{X} defined in Section VI-A, then the natural way to set ℓ and θ is to take $\ell(x, x') = \mathbf{1}_{\{x\}}(x')$ for all $x, x' \in \mathbf{X}$ and to define θ on the diagonal of $\mathbf{X} \times \mathbf{X}$ only, by $\theta(x, x) = x$ for all $x \in \mathbf{X}$. Since θ is bijective in this case, we can use (22) to find that

$$f \odot f' : x \mapsto \ell(x, x) f(x) f'(x) = f(x) f'(x).$$

In other words, it holds that $f \odot f' = f \cdot f'$ and the result of Section VI-A about fusion for state spaces is recovered. This confirms that the fusion operation introduced in this section is not different from the one of Section VI-A, it is instead a more general formulation of the same operation.

Example 14. If $\mathbf{Y} = \mathbf{C}_1(\mathbf{X})$ where \mathbf{X} is the state space defined in Section VI-A, then the natural way to set ℓ and θ is $\ell(P, P') = \|P * P'\|$ and $\theta(P, P') = P * P'$ for any probabilistic constraints P and P' in $\mathbf{C}_1(\mathbf{X})$. In this case, we find that

$$f \odot f' : \hat{P} \mapsto \sup_{(P, P') : P * P' = \hat{P}} \|P * P'\| f(P) f'(P').$$

The binary operation \odot is associative since it indeed holds that

$$\begin{aligned} (P * P') * P'' &= P * (P' * P'') \\ \|P * P'\| \| (P * P') * P'' \| &= \|P * (P' * P'')\| \|P' * P''\| \end{aligned}$$

for any $P, P', P'' \in \mathbf{C}_1(\mathbf{X})$. The set $\mathbf{C}_1(\mathbf{X})$ is one of the useful examples of sets on which the fusion operation takes a

general form since two probabilistic constraints do not have to be equal to represent the same individual. Also, the mapping θ is surjective but not bijective in general for probabilistic constraints, as opposed to the case of Example 13. This level of generality was required in [21, Chapt. 2] when deriving a principled solution to the problem of data association for multi-object dynamical systems since these systems are best represented by hierarchical level of uncertainty, as already identified in [26, 8].

CONCLUSION

Supremum-based outer measures have demonstrated their ability to represent uncertainty in a flexible way and to comply with intuitively appealing operations such as pullback and data assimilation. Future work includes the generalisation of the proposed type of outer measure to product spaces, where correlations between the different components of the product space can take a more involved form.

REFERENCES

- [1] J. Baron. Second-order probabilities and belief functions. *Theory and Decision*, 23(1):25–36, 1987.
- [2] L. E. Baum and T. Petrie. Statistical inference for probabilistic functions of finite state Markov chains. *The annals of mathematical statistics*, 37(6):1554–1563, 1966.
- [3] S. S. Blackman. *Multiple-target tracking with radar applications*, volume 1. Artech House, Norwood, 1986.
- [4] V. I. Bogachev. *Measure Theory*. Springer, 2007.
- [5] A. K. Brodzik and R. H. Enders. Semigroup structure of singleton Dempster-Shafer evidence accumulation. *IEEE transactions on Information theory*, 55(11):5241–5250, 2009.
- [6] H. Crauel. *Random probability measures on Polish spaces*, volume 11. CRC Press, 2003.
- [7] P. Del Moral. *Mean field simulation for Monte Carlo integration*. Chapman & Hall/CRC Monographs on Statistics & Applied Probability, 2013.
- [8] P. Del Moral and J. Houssineau. Particle association measures and multiple target tracking. In *Theoretical Aspects of Spatial-Temporal Modeling*, pages 1–30. Springer, 2015.
- [9] E. Delande, J. Houssineau, and D. E. Clark. Multi-object filtering with stochastic populations. *ArXiv preprint arXiv:1501.04671v2*, 2016.
- [10] A. P. Dempster. Upper and lower probability inferences based on a sample from a finite univariate population. *Biometrika*, 54(3-4):515–528, 1967.
- [11] A. P. Dempster. A generalization of Bayesian inference. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 205–247, 1968.
- [12] D. Dubois and H. Prade. *Possibility theory*. Wiley Online Library, 1988.
- [13] D. Dubois and H. Prade. Possibility theory: qualitative and quantitative aspects. In *Quantified representation of uncertainty and imprecision*, pages 169–226. Springer, 1998.

- [14] D. Dubois and H. Prade. *Fundamentals of fuzzy sets*, volume 7. Springer Science & Business Media, 2000.
- [15] R. Fagin, J. Y. Halpern, and N. Megiddo. A logic for reasoning about probabilities. *Information and computation*, 87(1):78–128, 1990.
- [16] T. S. Ferguson. A Bayesian analysis of some nonparametric problems. *The annals of statistics*, pages 209–230, 1973.
- [17] D. H. Fremlin. *Measure theory*. 4th edition, 2000.
- [18] N. Friedman and J. Y. Halpern. Plausibility measures and default reasoning. *Journal of the ACM*, 48(4):648–685, 2001.
- [19] J. Houssineau. *Representation and estimation of stochastic populations*. PhD thesis, Heriot-Watt University, 2015.
- [20] J. Houssineau and D. E. Clark. On a representation of partially-distinguishable populations. *arXiv preprint arXiv:1608.00723*, 2016.
- [21] J. Houssineau, D. E. Clark, and P. Del Moral. A sequential Monte Carlo approximation of the HISP filter. In *European Signal Processing Conference (EUSIPCO)*, 2015.
- [22] M. Loève. Probability theory, vol. II. *Graduate texts in mathematics*, 46:0–387, 1978.
- [23] R. P. S. Mahler. Can the Bayesian and Dempster-Shafer approaches be reconciled? Yes. In *IEEE International Conference on Information Fusion*, 2005.
- [24] J. Nedoma. Note on generalized random variables. In *Transactions of the First Prague Conference on Information Theory, Statistical Decision Functions, Random Processes (Liblice, 1956)*, Publishing House of the Czechoslovak Academy of Sciences, Prague, pages 139–141, 1957.
- [25] C. V. Negoita, L. A. Zadeh, and H. J. Zimmermann. Fuzzy sets as a basis for a theory of possibility. *Fuzzy sets and systems*, 1:3–28, 1978.
- [26] M. Pace and P. Del Moral. Mean-field PHD filters based on generalized Feynman-Kac flow. *IEEE Journal of Selected Topics in Signal Processing*, 7(3):484–495, 2013.
- [27] E. Schechter. *Handbook of Analysis and its Foundations*. Academic Press, 1996.
- [28] G. Shafer. *A mathematical theory of evidence*, volume 1. Princeton university press Princeton, 1976.
- [29] G. Shafer. The combination of evidence. *International Journal of Intelligent Systems*, 1(3):155–179, 1986.
- [30] R. C. Williamson. *Probabilistic arithmetic*. PhD thesis, University of Queensland, 1989.
- [31] J. Wishart. The generalised product moment distribution in samples from a normal multivariate population. *Biometrika*, pages 32–52, 1928.
- [32] J. Yen. Generalizing the Dempster-Schafer theory to fuzzy sets. *IEEE Transactions on Systems, Man and Cybernetics*, 20(3):559–570, 1990.
- [33] L. A. Zadeh. Fuzzy sets. *Information and Control*, 8:338–353, 1965.